

# Designing ISP-friendly Peer-to-Peer Networks Using Game-based Control

Vinith Reddy\*, Younghoon Kim<sup>†</sup>, Srinivas Shakkottai\* and A.L.Narasimha Reddy\*

\*Dept. of ECE, Texas A&M University

Email: {vinith\_reddy, sshakkot, reddy}@tamu.edu

<sup>†</sup>Dept. of CS, Korea Advanced Institute of Science and Technology

Email: kyhoon@gmail.com

**Abstract**—The rapid growth of peer-to-peer (P2P) networks in the past few years has brought with it increases in transit cost to Internet Service Providers (ISPs), as peers exchange large amounts of traffic across ISP boundaries. This ISP oblivious behavior has resulted in misalignment of incentives between P2P networks—that seek to maximize user quality—and ISPs—that would seek to minimize costs. Can we design a P2P overlay that accounts for both ISP costs as well as quality of service, and attains a desired tradeoff between the two? We design a system, which we call MultiTrack, that consists of an overlay of multiple *mTrackers* whose purpose is to align these goals. *mTrackers* split demand from users among different ISP domains while trying to minimize their individual costs (delay plus transit cost) in their ISP domain. We design the signals in this overlay of *mTrackers* in such a way that potentially competitive individual optimization goals are aligned across the *mTrackers*. The *mTrackers* are also capable of doing admission control in order to ensure that users who are from different ISP domains have a fair chance of being admitted into the system, while keeping costs in check. We prove analytically that our system is stable and achieves maximum utility with minimum cost. Our design decisions and control algorithms are validated by Matlab and ns-2 simulations.

## I. INTRODUCTION

The past few years have seen the rapid growth of content distribution over the Internet, particularly using peer-to-peer (P2P) networks. Recent studies estimate that 35-90% of bandwidth is consumed by P2P file-sharing applications, both at the edges and even within the core [1], [2]. The use of P2P networks for media delivery is expected to grow still further, with the proliferation of legal applications (e.g. Pando Networks [3]) that use P2P as a core technology.

While most P2P systems today possess some form of network resource-awareness, and attempt to optimally utilize the system resources, they are largely agnostic to Internet Service Providers' (ISP) concerns such as traffic management and costs. This ISP-oblivious nature of P2P networks has hampered the ability of system participants to correctly align incentives. Indeed, the recent conflicts between ISPs and content providers, as well as efforts by some ISPs such as Comcast to limit P2P traffic on their networks [4], speak in part to an inability to align interests correctly. Such conflicts are particularly critical as P2P becomes an increasingly prevalent form of content distribution [5].

A traditional BitTorrent system [6] has elements called *Trackers* whose main purpose is to enable peers to find

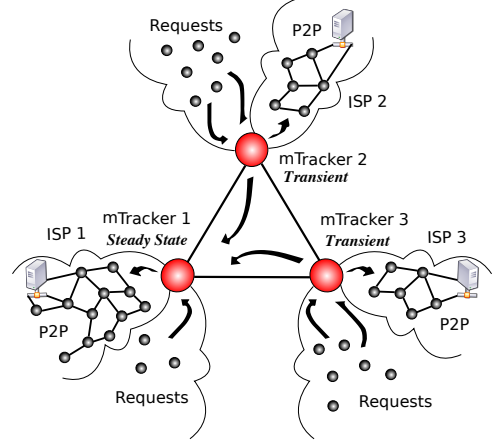


Fig. 1. The MultiTrack architecture. Multiple trackers, each following individual optimizations, achieve an optimal delay-cost tradeoff.

each other. The BitTorrent Tracker randomly assigns a new (entering) user a set of peers that are already in the system to communicate with. This system has the disadvantage that if peers who are assigned to help each other are in the domains of different ISPs, they would cause significant transit costs to the ISPs due to the inter-ISP traffic that they generate. However, if costs are reduced by forcing traffic to be local, then the delay performance of the system could suffer. Recent work such as [7]–[9] has focused on cost in terms of load balancing and localizing traffic, and developed heuristics to attain a certain quality of service (QoS). For example, P4P [8] develops a framework to achieve minimum cost (optimal load balancing) among ISP links, but its BitTorrent implementation utilizes the heuristic that 30% of peers declared to each requesting user should be drawn from “far away ISPs” in order to attain a good QoS.

This leads us to the fundamental question that we attempt to answer in this paper: *Can we develop a distributed delay and cost optimal P2P architecture?* We focus on developing a provably optimal price-assisted architecture called MultiTrack, that would be aware of the interaction between delay and cost. The idea is to understand that while the resources available with peers in different ISP domains should certainly be used, such usage comes at a price. The system must be able to determine the marginal gain in performance for a marginal increase in cost. It would then be able to locate the optimal point at which to operate.

The conceptual system<sup>1</sup> is illustrated in Figure 1. The system is managed by a set of *mTrackers*. Each *mTracker* is associated with a particular ISP domain. The *mTrackers* are similar to the Trackers in BitTorrent [6], in that their main purpose is to enable peers to find each other. However, unlike BitTorrent, the *mTrackers* in MultiTrack form an overlay network among themselves. The purpose of the overlay network is to provide multi-dimensional actions to the *mTrackers*. In Figure 1, *mTracker* 1 is in steady state (wherein the demand on the *mTracker* is less than the available capacity [11]), which implies that it has spare capacity to serve requests from other *mTrackers*. Consider *mTracker* 2 which is in transient state (wherein the demand on the *mTracker* is more than the available capacity [11]). When a request arrives, it can either assign the requester to its own domain at essentially zero cost, or can forward the user to *mTracker* 1 and incur a cost for doing so. However, the delay incurred by forwarded users would be less as *mTracker* 1 has higher capacity. Thus, *mTracker* 2 can trade-off cost versus delay performance by forwarding some part of its demand.

Each *mTracker* uses *price assisted decision making* by utilizing dynamics that consider the marginal payoff of forwarding traffic to that of retaining traffic in the same domain as the *mTracker*. Several such rational dynamics have been developed in the field of *game theory* that studies the behavior of selfish users. We present our system model with its simplifying assumptions in Section III. We then design a system in which the actions of these *mTrackers*, each seeking to maximize their own payoffs, actually result in ensuring lowest cost of the system as a whole. The scheme, presented in Section IV, involves *implicit learning* of capacities through probing and backoff through a rational control scheme known as replicator dynamics [12], [13]. We present a game theoretic framework for our system in Section IV-A and show using Lyapunov techniques that the vector of split probabilities converges to a provably optimal state wherein the total cost in terms of delay and traffic-exchange is minimized. Further, this state is a Wardrop equilibrium [14].

We then consider a subsidiary problem of achieving fair division of resources among different *mTrackers* through admission control in Section V. The objective here is to ensure that some level of fairness is maintained among the users in different *mTracker* domains, while at the same time ensuring that the costs in the system are not too high. Admission control implies that not all users in all domains would be allowed to enter the system, but it should be implemented in a manner that is fair to users in different *mTracker* domains. The *mTracker* takes admission control decisions based on the marginal disutility caused by users to the system. Users interested in the file would approach the *mTracker* that would decide whether or not to admit the user into the system. We show that our *mTracker* admission control optimally achieves fairness amongst users, while maintaining low system cost. Note that switching off admission control would still imply

that the total system cost is minimized by *mTrackers*, but this could be high if the offered load were high.

We simulate our system both using Matlab simulations in Section VI to validate our analysis, as well as ns-2 simulations in Section VII to show a plausible implementation of the system as a whole. The simulations strongly support our architectural decisions. We conclude with ideas on the future in Section VIII.

## II. RELATED WORK

There has been much recent work on P2P systems and traffic management, and we provide a discussion of work that is closely related to our problem. Fluid models of P2P systems, and the multi-phase (transient/steady state) behavior has been developed in [11], [15]. The results show how supply of a file correlates with its demand, and it is essentially transient delays that dominate.

Traffic management and load balancing have become important as P2P networks grow in size. There has been work on traffic management for streaming traffic [16]–[18]. In particular, [16] focuses on server-assisted streaming, while [17], [18] aim at fair resource allocation to peers using optimization-decomposition.

Closest to our setting is work such as [7]–[9], that study the need to localize traffic within ISP domains. In [7], the focus is on allowing only local communications and optimizing the performance by careful peer selection, while [8] develops an optimization framework to balance load across ISPs using cost information. A different approach is taken in [9], wherein peers are selected based on inputs on nearness provided by CDNs (if a CDN directs two peers to the same cache, they are probably near by).

Pricing and market mechanisms for P2P systems are of significant interest, and work such as [19] use ideas of currency exchange between peers that can be used to facilitate file transfers. The system we design uses prices between *mTrackers* that map to real-world costs of traffic exchange, but do not have currency exchanges between peers which still use BitTorrent style bilateral barter.

## III. THE MULTITRACK SYSTEM

MultiTrack is a hybrid P2P network architecture similar to BitTorrent [6], [20] in many ways, and we first review some control elements of BitTorrent. In the BitTorrent architecture a file is divided into multiple chunks, and there exists at least one *Tracker* for each file that keeps track of peers that contain the file in its entirety (such peers are called *seeds*) or some chunks of it (such peers are called *downloaders*). A new peer that wants to download a file needs to first locate a Tracker corresponding to the file. Information about Trackers for a file (among other information) is contained in .torrent files, which are hosted at free servers. Thus, the peer downloads the .torrent file, and locates a Tracker using this file.

When a peer sends a request to a Tracker corresponding to the file it wants, the Tracker returns the addresses of a set of peers (*seeds and downloaders*) that the new peer should

<sup>1</sup>We presented some basic ideas on the system as a poster [10].

contact in order to download the file. The peer then connects to a subset of the given peers and downloads chunks of the file from them. While downloading the file, a peer sends updates to the Tracker about its download status (number of chunks uploaded and downloaded). Hence, a tracker knows about the state of each peer that is present in its peer cloud (or *swarm*).

The MultiTrack architecture consists of BitTorrent-like trackers, which we call **mTrackers**. We associate one or more mTrackers to each ISP, with each mTracker controlling access to its own peer cloud. Note that all these mTrackers are identified with the *same* file. Unlike BitTorrent Trackers, mTrackers are aware of each other and form an overlay network among themselves. Each mTracker consists of two different modules:

- 1) **Admission control:** Unlike the BitTorrent tracker which has no control over admission decisions of peers, the mTracker can decide whether or not to admit a particular peer into the system. Once admitted, the peer is either served locally or is forwarded to a different mTracker based on the decision taken by the mTracker.
- 2) **Traffic management:** This module of the mTracker, takes a decision on whether to forward a new peer into its own peer cloud (at relatively low cost, but possibly poor delay performance) or to another mTracker (at higher cost, but potentially higher performance).

The rationale behind this architecture is as follows. At any time, a peer cloud has a capacity associated with it, based on the maximum upload bandwidth of a peer in the cloud and the total number of chunks present at all the peers in the cloud (seeds and downloaders). In general, a peer-cloud has two phases of operation [11]—a *transient* phase where the available capacity is less than the demand (in other words, not enough peers with a copy of the file), and a *steady state* phase, where the available capacity is greater than the capacity required to satisfy demand. Thus, a peer cloud can be thought of as a server with changing service capacity. We balance load among different peer clouds located in different ISPs, taking into account the transit cost associated with traffic exchange.

We assume *time scale separation* between the two modes—traffic management and admission control, of the mTracker. Our assumption is that the capacity of a P2P system remains roughly constant over intervals of time, with capacity changes seen at the end of these time periods. We divide system dynamics into three time scales:

- 1) **Large:** The capacity of the peer cloud associated with each mTracker changes at this time scale.
- 2) **Medium:** mTrackers take admission control decisions at this time scale. They could increase or decrease the number of admitted peers based on feedback from the system. We will study dynamics at this time scale in Section V.
- 3) **Small:** mTrackers split the demand that they see among the different options (mTrackers visible to them) at this time scale. Thus, they change the probability of sending peers to their own peer-cloud or to other mTrackers at

this time scale. We study these dynamics in Section IV. Note that a medium time unit comprises of many small time units and a large time unit comprises of many medium time units. The artifice of splitting dynamics into these time scales allows us to design each control loop while assuming that certain system parameters remain constant during the interval. In the following sections, we present the design and analysis of our different system components.

#### IV. MTRACKER: TRAFFIC MANAGEMENT

The objective of the mTracker's *Traffic Management* module is to split the demand that it sees among the different options (other mTrackers, and its own peer cloud) that it has. Since each mTracker is associated with a different ISP domain, it would like to minimize the cost seen by that ISP, and yet maintain a good delay performance for its users.

As mentioned in the last section, peer-clouds can be in either *transient* or *steady-state* based on whether the demand seen is greater than or less than the available capacity. An mTracker in the transient mode would like to offload some of its demand, while mTrackers in the steady-state mode can accept load. Thus, each mTracker  $j$  in the transient mode maintains a split probability vector  $\hat{y}_j = [\hat{y}_j^1 \dots \hat{y}_j^Q]$ , where  $Q$  is the total number of mTrackers, and some of the  $y_j^i$  could be zero. We assume that the demand seen by mTracker  $j$  is a Poisson process of rate  $x_j$ . Thus, splitting traffic according to  $\hat{y}_j$  would produce  $Q$  Poisson processes, each with rate  $x_j^i \triangleq y_j^i x_j$  ( $i = 1, \dots, Q$ ).

Now, each mTracker in the steady-state mode can accept traffic from mTrackers that are transient. It could, of course, prioritize or reserve capacity for its own traffic; we assume here that it does so, and the balance capacity available (in users served per unit time) of this steady state mTracker is  $C^i$ . Then the demand seen at each such mTracker  $i$  is the sum of Poisson processes that arrive at it, whose rate is simply  $\sum_{l=1}^Q x_l^i$ . We assume that delay seen by each peer sent to mTracker  $i$  is convex increasing in load, and for illustration use the M/M/1 delay function

$$\frac{1}{C^i - \sum_{l=1}^Q x_l^i}. \quad (1)$$

Note that, peers from different transient mTrackers are not allowed to communicate with each other at the steady state mTracker to which they are forwarded. Thus, a peer that is forwarded from one ISP domain to another is only allowed to communicate with peers located in that ISP domain.

Now, the steady state mTrackers are disinterested players in the system, and would like to minimize the total delay of the system. They could charge an additional price that would act as a congestion signal to mTrackers that forward traffic to them. Such a congestion price should reflect the ill-effect that increasing the load by one mTracker has on the others. What should such a price look like? Now, consider the expression

$$D(z) = \frac{1}{C^i - z^i}, \quad (2)$$

which is the general form of the delay seen by each user at mTracker  $i$ . The elasticity of delay with arrival rate  $z_i$

$$\frac{\partial D(z^i)}{\partial z^i} \frac{z^i}{D(z^i)} = \frac{z^i}{C^i - z^i}. \quad (3)$$

The elasticity gives the fractional change in delay for a fractional change in load, and can be thought of as the cost of increasing load on the users. In other words, if the load is increased by any one mTracker, *all* the others would also be hurt by this quantity. Expressing the above in terms of delay (multiplying by total delay) to ensure that all units are in delay, the elasticity per unit rate per unit time at mTracker  $i$  is just

$$\frac{\sum_{l=1}^Q x_l^i}{(C^i - \sum_{l=1}^Q x_l^i)^2}. \quad (4)$$

The above quantity represents the ill effect that increasing the load per unit time has on the delay experienced on all users at mTracker  $i$ . The delay cost (1) is the disutility for using the mTracker, while the congestion cost (4) is the disutility caused to others using the mTracker. The mTracker can charge this price to each mTracker that forward peers to it.

Since mTrackers belong to different ISP domains, forwarding demand from one mTracker to the other is not free. We assume that the transit cost per unit rate of forwarding demand from mTracker  $j$  to mTracker  $i$  is  $p_j^i$ . Thus, the payoff of mTracker  $j$  due to forwarding traffic to mTracker  $i$  per unit rate per unit time is given by the sum of transit cost  $p_j^i$  with the delay cost (1) and congestion price (4), which yields a total payoff per unit rate per unit time of

$$\frac{1}{C^i - \sum_{l=1}^Q x_l^i} + p_j^i + \frac{\sum_{l=1}^Q x_l^i}{(C^i - \sum_{l=1}^Q x_l^i)^2}. \quad (5)$$

The mTracker would like as *small* a payoff as possible.

In the next subsections we will develop a population game model for our system, and show how rational dynamics when coupled with the payoff function given above naturally results in minimizing the total system cost (delay cost plus transit cost). A good reference on population games is [21].

#### A. MultiTrack Population Game

A *population game*  $\mathcal{G}$ , with a set  $\mathcal{Q} = \{1, \dots, Q\}$  of non-atomic populations of players is defined by the following entities:

- 1) a mass,  $x_j \quad \forall j \in \mathcal{Q}$ ,
- 2) a strategy or action set,  $\mathcal{S}_j = \{1, \dots, S_j\} \quad \forall j \in \mathcal{Q}$  and
- 3) a payoff,  $F_j^i \quad \forall j \in \mathcal{Q}$  and  $\forall i \in \mathcal{S}_j$ .

By a non-atomic population, we mean that the contribution of each member of the population is infinitesimal.

In the *MultiTrack Game* each mTracker is a player and the options available to each mTracker are other mTrackers' peer cloud or its own peer cloud. Let  $\vec{x} = [x_1, \dots, x_Q]$  be the total load vector of the system at the small time scale, where  $x_j \quad \forall j \in \mathcal{Q}$  is the total arrival rate of new peer requests (or mass) at mTracker  $j$ . A strategy distribution of an mTracker  $j \in \mathcal{Q}$  is a split of its load  $x_j$  among different

mTrackers including itself, represented as  $\vec{x}_j = [x_j^1 \dots x_j^Q]$ , where  $\sum_{i=1}^Q x_j^i = x_j$ . If a mTracker  $j$  is not connected to mTracker  $i$  (or if it does not want to use mTracker  $i$ ), then the rate  $x_j^i = 0$ . We denote the vector of strategies being used by all the mTrackers as  $\mathbf{X} = [\vec{x}_1 \dots \vec{x}_Q]$ . The vector  $\mathbf{X}$  represents the state of the system and it changes continuously with time. Let the space of all possible states of a system for a given load vector be denoted as  $\mathbb{X}$ , i.e.  $\mathbf{X} \in \mathbb{X}$ .

The payoff (per unit rate per unit time) of forwarding requests from mTracker  $j$  to  $i$ , when the state of the system is  $\mathbf{X}$  is denoted by  $F_j^i(\mathbf{X}) \in \mathbb{R}$  and is assumed to be continuous and differentiable. As developed above this payoff is

$$F_j^i(\mathbf{X}) = \frac{1}{C^i - \sum_{l=1}^Q x_l^i} + p_j^i + \frac{\sum_{l=1}^Q x_l^i}{(C^i - \sum_{l=1}^Q x_l^i)^2}. \quad (6)$$

Recall that mTrackers want to keep payoff *as small as possible*.

A commonly used concept in non-cooperative games in the context of infinitesimal players, is the Wardrop equilibrium [14]. Consider any strategy distribution  $\vec{x}_j = [x_j^1, \dots, x_j^{S_j}]$ . There would be some elements which are non-zero and others which are zero. We call the strategies corresponding to the non-zero elements as the *strategies used by population  $j$* .

**Definition 1** A state  $\hat{\mathbf{X}}$  is a Wardrop equilibrium if for any population  $j \in \mathcal{Q}$ , all strategies being used by the members of  $j$  yield the same marginal payoff to each member of  $j$ , whereas the marginal payoff that would be obtained by members of  $j$  is higher for all strategies not used by population  $j$ .

In the context of our MultiTrack game the above definition of Wardrop equilibrium is characterized by the following relation:

$$F_j^r(\hat{\mathbf{X}}) \leq F_j^i(\hat{\mathbf{X}}) \quad \forall r \in \hat{\mathcal{Q}}_j \text{ and } i \in \mathcal{Q}$$

Where  $\hat{\mathcal{Q}}_j \subset \mathcal{Q}$  is the set of all mTrackers used by population  $j$  in a strategy distribution  $\hat{\mathbf{X}}$ .

The above concept refers to an *equilibrium condition*; the question arises as to how the system actually arrives at such a state. One model of population dynamics is *Replicator Dynamics* [12]. The rate of increase of  $\dot{x}_j^i/x_j^i$  of the strategy  $i$  is a measure of its evolutionary success. Following ideas of Darwinism, we may express this success as the difference in fitness  $F_j^i(\mathbf{X})$  of the strategy  $i$  and the average fitness  $\sum_{r=1}^Q x_j^r F_j^r(\mathbf{X})/x_j$  of the population  $j$ . Then we obtain

$$\frac{\dot{x}_j^i}{x_j^i} = \text{average fitness} - \text{fitness of } s.$$

Then the dynamics used to describe changes in the mass of population  $j$  playing strategy  $s$  is given by

$$\dot{x}_j^i = x_j^i \left( \frac{1}{x_j} \sum_{r=1}^Q x_j^r F_j^r(\mathbf{X}) - F_j^i(\mathbf{X}) \right). \quad (7)$$

The above expression implies that a population would increase the mass of a successful strategy and decrease the mass of a less successful one. It is called the replicator equation after the tenet "like begets like". Note that the total mass of the

population  $j$  is constant. We design our mTracker Traffic Management module around Replicator Dynamics (7).

### B. Convergence of mTracker dynamics

We define the total cost in the system to be the sum of the total system delay plus the total transit cost. In other words, we have weighted delay costs and transit costs equally when determining their contribution to the system cost. We could, of course, use any convex combination of the two without any changes to the system design. Hence using the M/M/1 delay model at each tracker and adding transit costs, the total system cost when the system is in state  $\mathbf{X}$  is given as:

$$\mathcal{C}(\mathbf{X}) = \sum_{i=1}^Q \left\{ \frac{\sum_{r=1}^Q x_r^i}{C^i - \sum_{l=1}^Q x_l^i} + \sum_{r=1}^Q p_r^i x_r^i \right\}. \quad (8)$$

Note that the cost is convex and increasing in the load. We will show that the above expression acts as a Lyapunov function for the system.

*Theorem 1:* The system of mTrackers that follow *replicator dynamics* with payoffs given by (6) is globally asymptotically stable.

*Proof:* We prove the system stability using Lyapunov Theory with  $\mathcal{C}(\mathbf{X})$  defined in (8) as the Lyapunov function. From (6) and (8),  $\frac{\partial \mathcal{C}}{\partial x_j^i} = F_j^i(\mathbf{X})$ , hence

$$\dot{\mathcal{C}}(\mathbf{X}) = \sum_{i=1}^Q \sum_{j=1}^Q \frac{\partial \mathcal{C}}{\partial x_j^i} \dot{x}_j^i = \sum_{i=1}^Q \sum_{j=1}^Q F_j^i(\mathbf{X}) \dot{x}_j^i, \quad (9)$$

Now, let  $\tilde{\mathcal{X}}$  be the set of states such that,

$$\dot{\mathcal{C}}(\mathbf{X}) = 0, \forall \mathbf{X} \in \tilde{\mathcal{X}}$$

From (9) it is evident that  $\dot{\mathcal{C}}(\mathbf{X}) = 0$ , if:

$$F_j^i(\mathbf{X}) = 0 \text{ or} \quad (10)$$

$$(\dot{x}_j^i = 0) \Rightarrow \frac{1}{x_j} \sum_{r=1}^Q x_j^r F_j^r = F_j^i \quad \forall i, j \in Q \quad (11)$$

Thus,  $\tilde{\mathcal{X}}$  is the set of equilibrium states of replicator dynamics. We will show that  $\dot{\mathcal{C}}(\mathbf{X}) < 0 \quad \forall \mathbf{X} \notin \tilde{\mathcal{X}}$ .

From (9) and (7) we have

$$\dot{\mathcal{C}}(\mathbf{X}) = \sum_{i=1}^Q \sum_{j=1}^Q F_j^i x_j^i \left( \frac{1}{x_j} \sum_{r=1}^Q x_j^r F_j^r - F_j^i \right) \quad (12)$$

$$= \sum_{j=1}^Q x_j \left( \left( \sum_{i=1}^Q \frac{x_j^i}{x_j} F_j^i \right)^2 - \left( \sum_{i=1}^Q \frac{x_j^i}{x_j} (F_j^i)^2 \right) \right) \quad (13)$$

Since function  $f(x) = x^2$  is convex and  $\sum_{i=1}^Q \frac{x_j^i}{x_j} = 1$ , from Jensen's inequality we have,  $\forall \mathbf{X} \notin \tilde{\mathcal{X}}$ :

$$\left( \left( \sum_{i=1}^Q \frac{x_j^i}{x_j} F_j^i \right)^2 - \left( \sum_{i=1}^Q \frac{x_j^i}{x_j} (F_j^i)^2 \right) \right) < 0 \quad \forall j \in Q$$

Thus,  $\dot{\mathcal{C}}(\mathbf{X}) < 0, \quad \forall \mathbf{X} \notin \tilde{\mathcal{X}}$  and the system is globally asymptotically stable [22]. ■

While replicator dynamics is a simple model, it has a drawback: during the different iterations of replicator dynamics, if the value of  $x_j^i$ , the rate of forwarding requests from mTracker  $j$  to mTracker  $i$  becomes zero then it remains zero forever. Thus, a strategy could become extinct when replicator dynamics is used and its stationary point might not be a Wardrop equilibrium. To avoid this problem, we can use another kind of dynamics called Brown-von Neuman-Nash (BNN) Dynamics, which is described as:

$$\dot{x}_q^i = x_q \gamma_q^i - x_q^i \sum_{j=1}^Q \gamma_q^j \quad (14)$$

$$\text{where, } \gamma_q^i = \max \left\{ F_q^i(X) - \frac{1}{x_q} \sum_{i=1}^Q x_q^i F_q^i(X), 0 \right\} \quad (15)$$

denotes the excess payoff to strategy  $i$  relative to the average payoff in population  $q$ . We can show that the system of mTrackers that follow BNN dynamics is globally asymptotically stable in a manner similar to the proof of Theorem 1.

We have just shown that the total system cost acts as a Lyapunov function for the system. It should not come as a surprise then, that the cost-minimizing state is a Wardrop equilibrium. We prove this formally in the next subsection.

### C. Cost efficiency of mTrackers

In previous work on selfish routing (e.g. [23]), it was shown that the Wardrop equilibrium does not result in efficient system performance. This inefficiency is referred to as the *price of anarchy*, and it is primarily due to selfish user-strategies. However, work on population games [21] suggests that carefully devised price signals would indeed result in efficient equilibria. We show now that the Wardrop equilibrium attained by mTrackers is efficient for the system as a whole.

The objective of our system is to minimize the total cost for a given load vector  $\vec{x} = [x_1, \dots, x_Q]$ . Here the total cost in the system is  $\mathcal{C}(\mathbf{X})$  and is defined in (8). This can be represented as the following constrained minimization problem:

$$\min_{\mathbf{X}} \quad \mathcal{C}(\mathbf{X}) \quad (16)$$

$$\text{subject to, } \sum_{i=1}^Q x_j^i = x_j \quad \forall j \in Q \quad (17)$$

$$x_j^i \geq 0 \quad (18)$$

The Lagrange dual associated with the above is

$$\mathcal{L}(\lambda, \mathbf{X}) = \max_{\lambda, h} \min_{\mathbf{X}} \left( \mathcal{C}(\mathbf{X}) - \right. \quad (19)$$

$$\left. \sum_{j=1}^Q \lambda_j \left( \sum_{i=1}^Q x_j^i - x_j \right) - \sum_{i=1}^Q \sum_{j=1}^Q h_j^i x_j^i \right)$$

where  $h_j^i \geq 0$  and  $\lambda_j, \forall i, j \in Q$  are the dual variables. Now the above dual problem gives the following Karush-Kuhn-Tucker first order conditions:

$$\frac{\partial \mathcal{L}}{\partial x_j^i}(\lambda, \mathbf{X}^*) = 0 \quad \forall i, j \in Q \quad (20)$$

and

$$h_j^i x_j^{*i} = 0 \quad \forall i, j \in Q \quad (21)$$

where  $\mathbf{X}^*$  is the global minimum for the primal problem (16). Hence, from (20) we have

$$\begin{aligned} \frac{\partial \mathcal{C}}{\partial x_j^i}(\mathbf{X}^*) - \lambda_j \frac{\partial(\sum_{i=1}^Q x_j^{*i} - x_j^*)}{\partial x_j^i} + h_j^i &= 0 \quad \forall i, j \in \mathcal{Q} \\ \Rightarrow \frac{\partial \mathcal{C}}{\partial x_j^i}(\mathbf{X}^*) &= \lambda_j + h_j^i \quad \forall i, j \in \mathcal{Q} \end{aligned} \quad (22)$$

We know from the definition of payoff (6) that  $\frac{\partial \mathcal{C}}{\partial x_j^i}(\mathbf{X}) = F_j^i(\mathbf{X})$ . Thus from (22) we have

$$F_j^i(\mathbf{X}^*) = \lambda_j + h_j^i \quad \forall i, j \in \mathcal{Q} \quad (23)$$

From (21), it follows that

$$F_j^i(\mathbf{X}^*) = \lambda_j \quad \text{when } x_j^{*i} > 0 \quad \forall i, j \in \mathcal{Q} \quad (24)$$

and

$$F_j^i(\mathbf{X}^*) = \lambda_j + h_j^i \quad \text{when } x_j^{*i} = 0 \quad \forall i, j \in \mathcal{Q} \quad (25)$$

Now, consider the replicator dynamics (7), at stationary point we have  $\dot{x}_j^i = 0$ . Thus,

$$\hat{F}_j = F_j^i(\hat{X}) \quad \forall i, j \in \mathcal{Q} \quad (26)$$

$$\text{or} \quad \hat{x}_j^i = 0,$$

where

$$\hat{F}_j \triangleq \frac{1}{\hat{x}_j} \sum_{r=1}^Q \hat{x}_j^r F_j^r(\hat{X}) \quad \forall j \in \mathcal{Q}, \quad (27)$$

and  $\hat{X}$  denotes a stationary point. The above equations imply that for mTracker  $j$  the per unit cost of forwarding traffic to all the other mTrackers that it uses is the same. However, for an option  $i$  that it does not use, the rate of forwarding  $x_j^i$  is 0 or equivalently, the cost is more than the average payoff. Finally, we observe that the conditions required for Wardrop equilibrium are identical to the KKT first order conditions (24)-(25) of the minimization problem (16) when

$$\hat{F}_j = \lambda_j \quad \forall j \in \mathcal{Q}, \quad (28)$$

which leads to the following theorem.

**Theorem 2:** The solution of the minimization problem in (16) is identical to the Wardrop equilibrium of the non-cooperative potential game  $\mathcal{G}$ .

*Proof:* From the above discussion we know that the KKT conditions of (16) satisfy the Wardrop equilibrium conditions of the game  $\mathcal{G}$ . Thus, to finish this proof all we need to show is that there is no duality gap between the primal (16) and the dual (19) problems. This follows immediately from convexity of the total system cost. ■

## V. MTRACKER: ADMISSION CONTROL

In the previous section we witnessed how each mTracker tries to reduce the cost in its peer cloud by forwarding requests to other mTrackers. However, minimizing the total delay does not mean that it is bounded. In order to ensure acceptable delay performance, we provide admission control functionality to each mTracker. The mTracker takes admission control decisions in the medium time scale; the mTracker's demand splitting is assumed to have converged to yield the lowest cost split at every instant at this time scale. In some

ways the admission control mode supplements natural market dynamics—if the delay experienced by requesters were unbearably high, they would simply abort, causing the system to recover. However, such dynamics might cause large swings in quality over time; the mTracker's admission control precludes the occurrence of such swings.

We could formulate an admission control problem, enforcing hard constraints on the acceptable system cost, as a convex optimization problem shown below:

$$\max_{\vec{x}} \quad \sum_{j=1}^Q w_j \log x_j \quad (29)$$

$$\text{subject to,} \quad \mathcal{C}^*(\vec{x}) \leq \kappa \quad (30)$$

$$x_j \geq 0$$

where  $\vec{x}$  is the load vector and  $\mathcal{C}^*(\vec{x})$  is the minimum value of the optimization problem (16) for a given load  $\vec{x}$ . We can easily show that the constraint set in (29) is convex.

**Lemma 3:** The set of all load vectors  $\vec{x}$ , satisfying the inequality constraint,  $\mathcal{C}^*(\vec{x}) \leq \kappa$  is convex.

*Proof:* Let  $\vec{x}$  and  $\vec{y}$  be two load vectors such that,

$$\mathcal{C}^*(\vec{x}) \leq \kappa \text{ and } \mathcal{C}^*(\vec{y}) \leq \kappa \quad (31)$$

Let  $X_{min}$  and  $Y_{min}$  be the states, corresponding to load vectors  $\vec{x}$  and  $\vec{y}$  respectively, which results in minimum cost to the system, i.e.,

$$\mathcal{C}(X_{min}) = \mathcal{C}^*(\vec{x}) \text{ and } \mathcal{C}(Y_{min}) = \mathcal{C}^*(\vec{y}) \quad (32)$$

Consider,

$$\mathcal{C}(\alpha X_{min} + (1 - \alpha)Y_{min}) \leq \alpha \mathcal{C}(X_{min}) + (1 - \alpha)\mathcal{C}(Y_{min}) \quad (33)$$

the above inequality follows from the convexity of  $\mathcal{C}(X)$ .

Using Eqns(31) and (32), we get:

$$\mathcal{C}(\alpha X_{min} + (1 - \alpha)Y_{min}) \leq \alpha \mathcal{C}^*(\vec{x}) + (1 - \alpha)\mathcal{C}^*(\vec{y}) \quad (34)$$

$$\leq \alpha \kappa + (1 - \alpha)\kappa \quad (35)$$

$$\leq \kappa \quad (36)$$

if  $\vec{z} = \alpha \vec{x} + (1 - \alpha)\vec{y}$ , then from the definition of  $\mathcal{C}^*$

$$\mathcal{C}(Z_{min}) = \mathcal{C}^*(\vec{z})\mathcal{C}^*(\vec{z}) = \mathcal{C}(Z_{min}) \quad (37)$$

where  $Z_{min}$  is the state of the system, corresponding to load  $\vec{z}$ , when the cost is minimum.

Clearly we can represent any state  $Z$ , corresponding to the load vector  $\vec{z}$ , in the form of  $\alpha X + (1 - \alpha)Y$ , and thus it follows from the definition of  $\mathcal{C}^*$  and Eqn(36) that:

$$\mathcal{C}^*(\alpha \vec{x} + (1 - \alpha)\vec{y}) \leq \mathcal{C}(\alpha X_{min} + (1 - \alpha)Y_{min}) \leq \kappa \quad (38)$$

Thus the set is convex. ■

If we think of  $\sum_j w_j \log x_j$  as the total system utility, then  $\mathcal{C}^*(\vec{x})$  is the total system disutility. Instead of hard constraints on the cost, we relax the problem after the manner of [24], [25] to simply ensure that the difference of utility and disutility (the *net utility*) is as large as possible.

$$\max_{\vec{x}} \quad \sum_{j=1}^Q w_j \log x_j - \mathcal{C}^*(\vec{x}) \quad (39)$$

$$\text{subject to,} \quad x_j \geq 0$$

A gradient ascent type controller that could be used to solve the above problem is

$$\dot{x}_j = w_j - x_j \frac{\partial \mathcal{C}^*}{\partial x_j} \forall j \in \mathcal{Q} \quad (40)$$

To determine the second term above, we use

$$\frac{\partial \mathcal{C}^*}{\partial x_j} = \sum_{i \in \mathcal{Q}} \frac{\partial \mathcal{C}(\mathbf{X}^*)}{\partial x_j^i} \frac{\partial x_j^i}{\partial x_j} + \sum_{k \in \mathcal{Q}, k \neq j} \sum_{i \in \mathcal{Q}} \frac{\partial \mathcal{C}(\mathbf{X}^*)}{\partial x_k^i} \frac{\partial x_k^i}{\partial x_j} \forall j \in \mathcal{Q} \quad (41)$$

$$= \sum_{i \in \mathcal{Q}} F_j^i(\mathbf{X}^*) \frac{\partial x_j^i}{\partial x_j} + \sum_{k \in \mathcal{Q}, k \neq j} \sum_{i \in \mathcal{Q}} F_k^i(\mathbf{X}^*) \frac{\partial x_k^i}{\partial x_j} \forall j \in \mathcal{Q} \quad (42)$$

When the system is in Wardrop equilibrium ( $\mathbf{X}^*$ ) all the options  $i$  used by  $j$  yield the same payoff hence,  $F_j^i(\mathbf{X}^*) = F_j^*$ . For the options  $r$  that were not used  $\frac{\partial x_j^r}{\partial x_j} = 0$  at state  $\mathbf{X}^*$ .

Further,  $\sum_{i \in \mathcal{Q}} x_j^i = x_j$  and hence  $\sum_{i \in \mathcal{Q}} \frac{\partial x_j^i}{\partial x_j} = 1$ . Also,  $\sum_{i \in \mathcal{Q}} \frac{\partial x_k^i}{\partial x_j} = 0 \forall k \neq j$ . Thus, we just have

$$\frac{\partial \mathcal{C}^*}{\partial x_j} = F_j^* \quad \forall j \in \mathcal{Q} \quad (43)$$

and the controller equation is given as:

$$\dot{x}_j = (w_j - x_j F_j^*) \quad \forall j \in \mathcal{Q}. \quad (44)$$

Under this admission control loop, we have the following theorem.

**Theorem 4:** Under the time scale separation assumption, the mTracker system with dynamics (44) is globally asymptotically stable.

*Proof:* We use the following Lyapunov function

$$Z(\vec{x}) = V(\hat{\vec{x}}) - V(\vec{x}) \quad (45)$$

$$\text{where } V(\vec{x}) = \sum_{j=1}^Q w_j \log x_j - \mathcal{C}^*(\vec{x}) \quad (46)$$

which is strictly concave, with  $\hat{\vec{x}}$  is its unique maximum. Differentiating  $Z(\vec{x})$  we obtain

$$\dot{Z} = - \sum_{j=1}^Q \frac{\partial V}{\partial x_j} \dot{x}_j. \quad (47)$$

Then from (46) and (44)

$$\frac{\partial V}{\partial x_j} = \frac{w_j}{x_j} - \frac{\partial \mathcal{C}^*(\vec{x})}{\partial x_j} = \frac{\dot{x}_j}{x_j} \quad (48)$$

$$\therefore \dot{Z} = - \sum_{j=1}^Q \frac{\dot{x}_j^2}{x_j} \leq 0 \quad \forall \vec{x}, \quad (49)$$

with  $\dot{Z} = 0$  when the system is in equilibrium. Thus, the system is globally asymptotically stable [22]. ■

Finally, we note that the equilibrium conditions of the controller (44) are the same as the KKT conditions of the convex optimization problem (39). Hence, the controller succeeds in maximizing the required net utility.

## VI. MATLAB SIMULATIONS

Figure 2 shows the per unit payoffs, corresponding to T2. As expected, the per unit payoffs converge to identical values. We

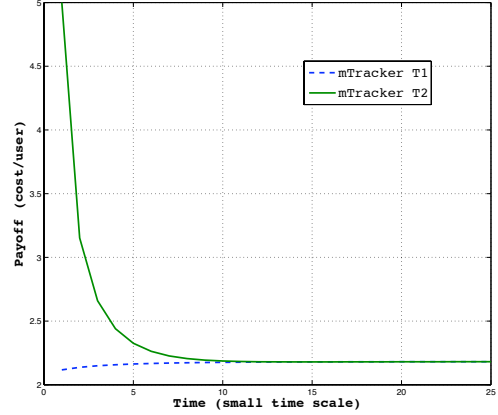


Fig. 2. The trajectory of payoffs of mTracker T2 for the 2 options available (local swarm and T1's swarm). The payoffs eventually equalize, showing that a Wardrop equilibrium has been attained.

perform simulations on the simple overlay topology illustrated in Figure 1. Our objective is to validate our analytical results, and use the resulting insights to construct a realistic ns-2 implementation in the next section. Our system consists of 3 mTrackers (T1, T2 and T3). The mTracker-T1 is assumed to be in steady state (i.e. it has more capacity than demand in its peer swarm) and the other mTrackers T2 and T3 are in a transient state. Thus, T2 and T3 may forward traffic to T1. Our simulation parameters are chosen as follows. The initial arrival rates at the mTrackers are  $x_1 = 10$  users/time,  $x_2 = 20$  users/time and  $x_3 = 20$  users/time, while the available capacities (fixed) are  $C^1 = 30$  users/time,  $C^2 = 20$  users/time and  $C^3 = 20$  users/time, respectively. There is a transit price for traffic forwarding between mTrackers, and these values are chosen as  $p_2^1 = 2$  unit and  $p_3^1 = 1$  unit.

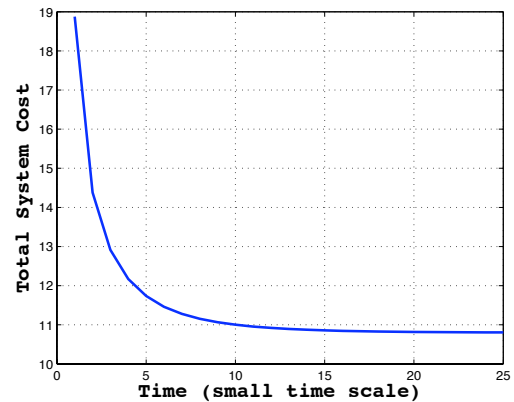


Fig. 3. The trajectory of total system cost in the system. As expected, it decreases over time to a minimum.

We first validate the dynamics of the mTrackers at the small time scale. Hence, the arrival rate at each mTracker remains fixed, and as in Section IV and they each use replicator dynamics in order to balance their payoffs among available



options. We expect that (i) the per unit payoff for all available options to an mTracker should eventually be equal, and (ii) the total cost of the system would decrease to a minimum. We observe similar convergence for mTracker T3. We then plot the trajectory of total system cost  $\mathcal{C}(\mathbf{X})^2$  in Figure 3. As expected it decreases with time, and converges to a minimum.

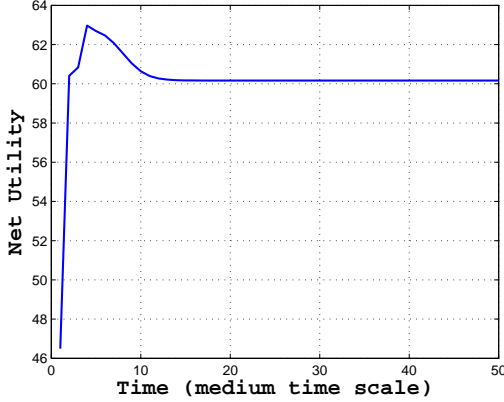


Fig. 4. The trajectory of net utility of the system when mTracker uses admission control. The net utility converges to a maximum.

We performed simulations at the medium time scale for the admission control module and observed that the net utility of the system (as defined in (39)) converges, shown in Figure 4.

While our Matlab simulations suggest that our system design is valid, they do not capture the true P2P interactions within each peer-cloud. In the next section, we implement MultiTrack using ns-2 in order to experiment with a more accurate representation of the system.

## VII. NS-2 EXPERIMENTS

We implemented the MultiTrack system on ns-2 to observe its performance in a more realistic setting. Again, we use the same network shown in Figure 1, with 3 mTrackers T1, T2 and T3. However, we now explicitly model peer behavior using a BitTorrent model. We use a flow level BitTorrent model developed by Eger *et al.*, [26]<sup>3</sup>, and each peer leaves the system after completing its download. We extended the existing BitTorrent Tracker model to support mTracker functionality.

We estimate the delay and congestion price at each mTracker during every small time scale as follows:

- 1) **Delay:** The per unit delay in each mTracker's peer cloud is measured by calculating the average download rate obtained by the peers in the current time slot, including the peers that finished service during this time slot. The delay experienced is the reciprocal of this download rate. We maintain an exponential moving average of the delay with 75% weight to the delay in current time slot and 25% weight to the previous value

<sup>2</sup>Recall that this is the sum of total delay plus total transit cost.

<sup>3</sup>Here only flows are simulated, and the actual dynamics of transport protocol, like TCP, and network protocol, like IP, are ignored in the interest of lowered simulation time.

- 2) **Congestion Price:** The congestion price with delay  $D$  and arrival rate  $z$  is given as  $\frac{\partial D}{\partial z} \times z$ , which follows from the *elasticity* (3). We measure the change in delay and change in arrival rate from the previous and current time slot to calculate the congestion price.

In our simulation, each peer has an upload capacity of 3000 kbps and their download capacity is not restricted. The requested file size is 5 MB and each chunk has a size of 256 kB. Peer arrivals are created according to Poisson processes of different rates. T1 has 200 seeds in its peer swarm while T2 and T3 have 5 seeds each. We fix the initial arrival rates to be  $x_1 = 3$  users/sec,  $x_2 = 5$  users/sec and  $x_3 = 7$  users/sec, set transit costs to be  $p_2^1 = 20$  and  $p_3^1 = 10$ .

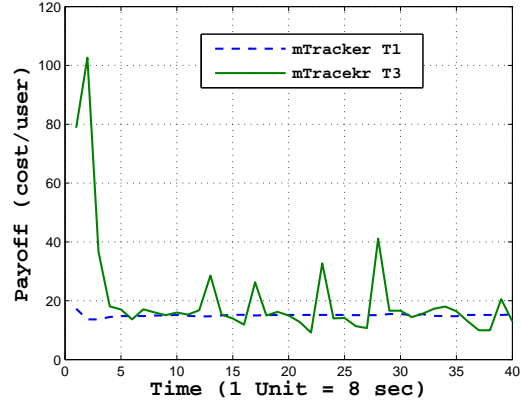


Fig. 5. The trajectory of payoffs of mTracker T3 for the 2 options available (local swarm and T1's swarm). The payoffs eventually equalize, showing that a Wardrop equilibrium has been attained.

We first simulate the mTracker with admission control disabled so as to show the convergence of our mTracker traffic management module. We set the update interval for the mTracker to be 8 sec. Thus, each mTracker calculates the splitting probabilities for the different options at this frequency. We simulate the system for 320 sec.

First we show the payoff convergence of the transient mTrackers. Figure 5, shows the convergence of payoffs of mTracker T3 for its two options, local swarm and T1's swarm, thus showing that the system attains Wardrop equilibrium. We observed similar payoff convergence for other mTrackers.

Next we plot the total cost (transit price and delay cost) of our MultiTrack system. The temporal evolution of cost is shown in Figure 6. The impact of using MultiTrack is clearly illustrated here. The system without traffic splitting has a high cost due to increased user delays, while traffic splitting without regard to prices has a high cost due to excessive transit traffic. MultiTrack takes both transit price and user delay into account, and hence achieves the lowest possible cost.

We implemented the mTracker's admission control module in ns-2. Here, at each time step the mTracker decides the admission rate (based on the dynamics developed in Section V). Admission control is done at 40 sec intervals, while the traffic management module is run at 8 sec intervals during this interval. We simulate the system for 2000 sec. We expect that the net utility of the system (as defined in (39)) would increase



to a maximum, which is what we observe in Figure 7.

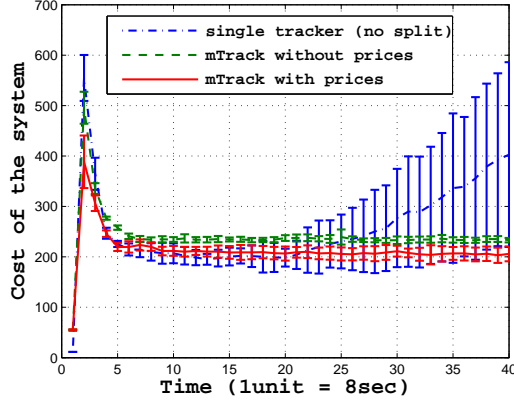


Fig. 6. The trajectory of total system cost. Without traffic splitting, the cost (delay plus transit price) is high. With traffic splitting without regard to price, the delay is low but transit price is high, causing high cost. MultiTrack takes prices and delays into account, and has lowest total cost.

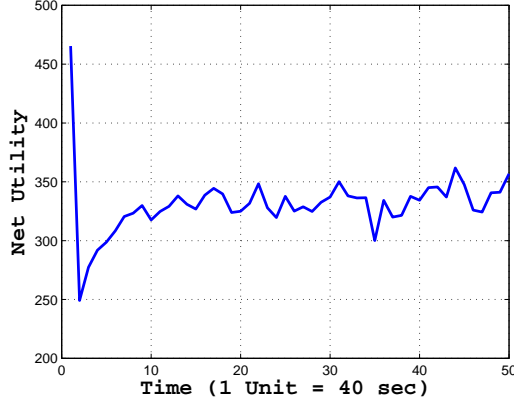


Fig. 7. The trajectory of net utility of the system when mTracker uses admission control.

In Figure 8 we can see the convergence of arrival rates of each mTracker, thus finding the optimum arrival rate into each mTracker for a fixed capacity. Since all mTrackers have identical utility, we see that T1 dominates as its price to access its own (resource rich) swarm is zero. Finally, we note that changing the time scales for faster responses does not seem to unduly impact stability. In particular, reducing the small time scale from 8 to 4 sec does not appreciably change our results.

## VIII. CONCLUSIONS

As the popularity of P2P systems has grown, it has become clear that aligning incentives between the system performance in terms of the user QoS, and the transit costs faced by ISPs will be increasingly important. Fundamental to this problem is the realization that resources may be distributed geographically, and hence the marginal performance gain obtained by accessing a resource is offset in part by the marginal cost of transit in accessing it. In this paper, we consider delay and transit costs as two dimensions and attempt to design a system—MultiTrack—that attains an optimal operating point.

Our system consists of mTrackers that form an overlay network among themselves and act as gateways to peer-clouds. The load balancing module takes decisions based on whether

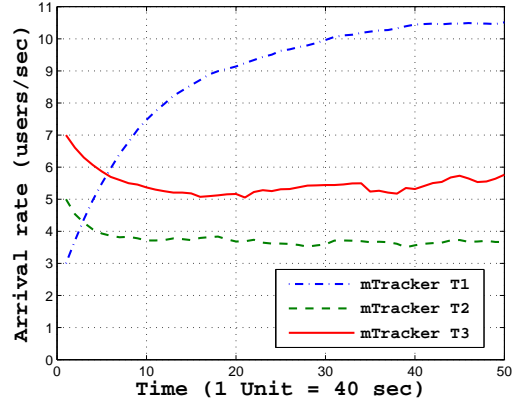


Fig. 8. The trajectory of arrival rates for all mTrackers. The utility of each mTracker is weighed equally to a value of 10.

the marginal decrease in delay obtained by forwarding a user to a resource rich peer-cloud is offset by the marginal increase in its transit cost. We show that a simple price-based controllers can ensure that the total system cost is minimized in spite of each mTracker being selfish. The admission control module calculates the tradeoff between the marginal utility in increasing the admission rate in a particular ISP domain to the marginal increase in system cost, to decide the admission rates into that ISP domain. It thus allows the correct arrival rate of users into the system to attain optimal performance.

We validated our system design using Matlab simulations, and implemented the system on ns-2 to conduct more realistic experiments. We showed that our system significantly outperforms a system in which costs are the only control dimension (localized traffic only). In the future, we will conduct testbed experiments on MultiTrack in a real-world setting.

## REFERENCES

- [1] C. Fraleigh, S. Moon, B. Lyle, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-level traffic measurements from the Sprint IP backbone," *IEEE Network*, vol. 17, no. 6, pp. 6–16, 2003.
- [2] K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan, "Measurement, modeling, and analysis of a peer-to-peer file-sharing workload," in *Proc. SOSP*, Oct 2003.
- [3] Pando Networks, Inc., 2008, <http://www.pando.com/>.
- [4] A. Broache, "FCC chief grills Comcast on BitTorrent blocking," *C|Net News.com*, Feb. 25, 2008.
- [5] E. Bangeman, "P2P responsible for as much as 90 percent of all 'Net traffic,'" *ArsTechnica*, Sept. 3, 2007.
- [6] B. Cohen, "Incentives to build robustness in BitTorrent," 2003, documentation from BitTorrent: <http://www.bittorrent.com>.
- [7] V. Aggarwal, A. Feldmann, and C. Scheidele, "Can ISPs and P2P users cooperate for improved performance?" *ACM Computer Communication Review*, vol. 37, no. 3, Jul 2007.
- [8] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz, "P4P: Portal for P2P applications," in *Proc. ACM SIGCOMM*, aug 2008.
- [9] D. R. Choffnes and F. Bustamante, "Taming the torrent: A practical approach to reducing cross-ISP traffic in p2p systems," in *Proc. ACM SIGCOMM*, Seattle, WA, Aug 2008.
- [10] V. Reddy, Y. Kim, S. Shakkottai, and A. L. N. Reddy, "MultiTrack: A Delay and Cost Aware P2P Overlay Architecture," in *ACM SIGMETRICS as a poster*, Seattle, WA, Jun. 2009.
- [11] X. Yang and G. de Veciana, "Performance of Peer-to-Peer Networks: Service Capacity and Role of Resource Sharing Policies," *Performance Evaluation: Special Issue on Performance Modeling and Evaluation of P2P Computing Systems*, vol. 63, 2006.
- [12] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

- [13] G. W. Brown and J. von Neumann, "Solution of games by differential equations," *Contributions to the Theory of Games I, Annals of Mathematical Studies*, vol. 24, 1950.
- [14] J. G. Wardrop, "Some theoretical aspects of road traffic research," in *Proc. of the Institute of Civil Engineers*, vol. 1, 1952, pp. 325–378.
- [15] D. Qiu and R. Srikant, "Modeling and performance analysis of BitTorrent-like peer-to-peer networks," in *Proc. ACM SIGCOMM*, Portland, Oregon, USA, August 2004.
- [16] E. Setton and J. Apostolopoulos, "Towards quality of service for peer-to-peer video multicast," in *Proc. ICIP*, Sep 2007.
- [17] S. Liu, R. Zhang-Shen, W. Jiang, J. Rexford, and M. Chiang, "Performance bounds for peer-assisted live streaming," in *Proc. ACM SIGMETRICS*, June 2008.
- [18] M. Chen, M. Ponec, S. Sengupta, J. Li, and P. A. Chou, "Utility maximization in peer-to-peer Systems," in *Proc. ACM SIGMETRICS*, June 2008.
- [19] C. Aperijs, M. J. Freedman, and R. Johari, "Peer-assisted content distribution with prices," in *Proc. ACM SIGCOMM CoNext*, Dec 2008.
- [20] "BitTorrent," <http://www.bittorrent.com/>, 2005.
- [21] W. H. Sandholm, "Potential Games with Continuous Player Sets," *Journal of Economic Theory*, vol. 97, pp. 81–108, January 2001.
- [22] H. Khalil, *Nonlinear Systems*. Prentice Hall, 1996.
- [23] T. Roughgarden and E. Tardos, "How bad is selfish routing?" in *IEEE Symposium on Foundations of Computer Science*, 2000, pp. 93–102.
- [24] F. P. Kelly, "Models for a self-managed Internet," *Philosophical Transactions of the Royal Society*, vol. A358, pp. 2335–2348, 2000.
- [25] S. Shakkottai and R. Srikant, *Network Optimization and Control*. Delft, The Netherlands: Now Publishes, 2008.
- [26] K. Eger, T. Hoßfeld, A. Binzenhöfer, and G. Kunzmann, "Efficient simulation of large-scale p2p networks: Packet-level vs. flow-level simulations," in *2nd Workshop on the Use of P2P, GRID and Agents for the Development of Content Networks (UPGRADE-CN'07)*, Monterey Bay, USA, jun 2007, pp. 9–16.